The use of machine learning with signal- and NLP processing of source code to detect and classify vulnerabilities and weaknesses with MARFCAT

Serguei A. Mokhov

Concordia University
Faculty of Engineering and Computer Science
Concordia University, Montréal, Québec, Canada
mokhov@cse.concordia.ca

SATE 2010

Introduction

- We present a machine learning approach to static code analysis for weaknesses related to security and others with the open-source MARF framework and its application to for the NIST's SATE 2010 static analysis tool exposition workshop [ODBN10].
- MARFCAT – MARF-based Code Analysis Tool [Mok10c]
- MARF [The10, Mok08, MD08, Mok10b, Mok10a]

Core principles

- Machine learning
- Spectral and NLP techniques

CVEs – the "Knowledge Base"

- Teach the system from the CVE-based cases
- Test on the CVE-based cases
- Test on the non-CVE-based cases

Categories for Machine Learning

- CVEs [NIS10a, NIS10b]
- CWEs [VM10] and/or our custom-made, e.g. per our classification methodology in [MLB07]
- Types (sink, path, fix)
- Line numbers (!)

Basic Methodology

- Compile meta-XML files from the CVE reports (line numbers, CVE, CWE, fragment size, etc.). Partly done by a Perl script and partly manually.
- Train the system based on the meta files to build the knowledge base (learn).
- Test on the training data for the same case (e.g. Tomcat 5.5.13 on 5.5.13 with the same annotations).
- Test on the testing data for the same case (e.g. Tomcat 5.5.13 on 5.5.13 without the annotations).
- Test on the testing data for the fixed case of the same software (Tomcat 5.5.13 on Tomcat 5.5.29).
- Test on the testing data for the general case (Tomcat 5.5.13 on Pebble).

Preliminary Results I

Current top precision:

- Wireshark:
  - CVEs (signal): 92.68%, CWEs (signal): 86.11%,
  - CVEs (NLP): 83.33%, CWEs (NLP): 58.33%
- Tomcat:
  - CVEs (signal): 83.72%, CWEs (signal): 81.82%,
  - CVEs (NLP): 87.88%, CWEs (NLP): 39.39%
- Chrome:
  - CVEs (signal): 90.91%, CWEs (signal): 100.00%,
- Dovecot:
  - 14 warnings; but it appears all quality or false positive
  - (very hard to follow the code, severely undocumented)
- Pebble:
  - none found during quick testing :-(

Preliminary Results II

- What follows are some select statistical measurements of the precision in recognizing CVEs and CWEs under different configurations using the signal processing and NLP processing. The complete set of statistics submitted is with the SATE-released data and even more complete is with the companion paper [Mok10c].
- "Second guess" statistics provided to see if the hypothesis that if our first estimate of a CVE/CWE is incorrect, the next one in line is probably the correct one. Both are counted if the first guess is correct.

Preliminary Results III

Wireshark, CVE-based

```
guess,run,config,good,bad,%
1st,1,-nopreprep -raw -fft -diff ,38,3,92.68
1st,2,-nopreprep -raw -fft -cheb ,38,3,92.68
1st,3,-nopreprep -raw -fft -eucl ,29,12,70.73
1st,4,-nopreprep -raw -fft -hamming ,26,15,63.41
1st,5,-nopreprep -raw -fft -mink ,23,18,56.10
1st,6,-nopreprep -raw -fft -cos ,37,51,42.05
2nd,1,-nopreprep -raw -fft -diff ,39,2,95.12
2nd,2,-nopreprep -raw -fft -cheb ,39,2,95.12
2nd,3,-nopreprep -raw -fft -eucl ,34,7,82.93
2nd,4,-nopreprep -raw -fft -hamming ,28,13,68.29
2nd,5,-nopreprep -raw -fft -mink ,31,10,75.61
2nd,6,-nopreprep -raw -fft -cos ,38,50,43.18
guess,run,config,good,bad,%
1st,1,CVE-2009-3829,6,0,100.00
1st,2,CVE-2009-2563,6,0,100.00
1st,3,CVE-2009-2562,6,0,100.00
1st,4,CVE-2009-4378,6,0,100.00
1st,5,CVE-2009-4376,6,0,100.00
```

Preliminary Results IV

```
1st,6,CVE-2010-0304,6,0,100.00
1st,7,CVE-2010-2286,6,0,100.00
1st,8,CVE-2010-2283,6,0,100.00
1st,9,CVE-2009-3551,6,0,100.00
1st,10,CVE-2009-3550,6,0,100.00
1st,11,CVE-2009-3549,6,0,100.00
1st,12,CVE-2009-3241,16,8,66.67
1st,13,CVE-2010-1455,34,20,62.96
1st,14,CVE-2009-3243,18,11,62.07
1st,15,CVE-2009-2560,8,6,57.14
1st,16,CVE-2009-2561,6,5,54.55
1st,17,CVE-2010-2285,6,5,54.55
1st,18,CVE-2009-2559,6,5,54.55
1st,19,CVE-2010-2287,6,6,50.00
1st,20,CVE-2009-4377,12,15,44.44
1st,21,CVE-2010-2284,6,9,40.00
1st,22,CVE-2009-3242,7,12,36.84
2nd,1,CVE-2009-3829,6,0,100.00
2nd,2,CVE-2009-2563,6,0,100.00
2nd,3,CVE-2009-2562,6,0,100.00
2nd,4,CVE-2009-4378,6,0,100.00
2nd,5,CVE-2009-4376,6,0,100.00
```

Preliminary Results V

```
2nd,6,CVE-2010-0304,6,0,100.00
2nd,7,CVE-2010-2286,6,0,100.00
2nd,8,CVE-2010-2283,6,0,100.00
2nd,9,CVE-2009-3551,6,0,100.00
2nd,10,CVE-2009-3550,6,0,100.00
2nd,11,CVE-2009-3549,6,0,100.00
2nd,12,CVE-2009-3241,17,7,70.83
2nd,13,CVE-2010-1455,44,10,81.48
2nd,14,CVE-2009-3243,18,11,62.07
2nd,15,CVE-2009-2560,9,5,64.29
2nd,16,CVE-2009-2561,6,5,54.55
2nd,17,CVE-2010-2285,6,5,54.55
2nd,18,CVE-2009-2559,6,5,54.55
2nd,19,CVE-2010-2287,12,0,100.00
2nd,20,CVE-2009-4377,12,15,44.44
2nd,21,CVE-2010-2284,6,9,40.00
2nd,22,CVE-2009-3242,7,12,36.84
```

Preliminary Results VI

Wireshark, CWE-based

```
guess,run,config,good,bad,%
1st,1,-cweid -nopreprep -raw -fft -cheb ,31,5,86.11
1st,2,-cweid -nopreprep -raw -fft -diff ,31,5,86.11
1st,3,-cweid -nopreprep -raw -fft -eucl ,29,7,80.56
1st,4,-cweid -nopreprep -raw -fft -hamming ,22,14,61.11
1st,5,-cweid -nopreprep -raw -fft -cos ,33,25,56.90
1st,6,-cweid -nopreprep -raw -fft -mink ,20,16,55.56
2nd,1,-cweid -nopreprep -raw -fft -cheb ,33,3,91.67
2nd,2,-cweid -nopreprep -raw -fft -diff ,33,3,91.67
2nd,3,-cweid -nopreprep -raw -fft -eucl ,33,3,91.67
2nd,4,-cweid -nopreprep -raw -fft -hamming ,27,9,75.00
2nd,5,-cweid -nopreprep -raw -fft -cos ,41,17,70.69
2nd,6,-cweid -nopreprep -raw -fft -mink ,22,14,61.11
guess,run,config,good,bad,%
1st,1,CWE-399,6,0,100.00
1st,2,NVD-CWE-Other,17,3,85.00
1st,3,CWE-20,50,10,83.33
1st,4,CWE-189,8,2,80.00
1st,5,NVD-CWE-noinfo,72,40,64.29
```

Preliminary Results VII

```
1st,6,CWE-119,13,17,43.33
2nd,1,CWE-399,6,0,100.00
2nd,2,NVD-CWE-Other,17,3,85.00
2nd,3,CWE-20,52,8,86.67
2nd,4,CWE-189,8,2,80.00
2nd,5,NVD-CWE-noinfo,83,29,74.11
2nd,6,CWE-119,23,7,76.67
```

Preliminary Results VIII

Wireshark, CVE-based (NLP)

```
guess,run,config,good,bad,%
1st,1,-noprepep -char -unigram -add-delta ,30,6,83.33
2nd,1,-noprepep -char -unigram -add-delta ,31,5,86.11
guess,run,config,good,bad,%
1st,1,CVE-2009-3829,1,0,100.00
1st,2,CVE-2009-2563,1,0,100.00
1st,3,CVE-2009-2562,1,0,100.00
1st,4,CVE-2009-4378,1,0,100.00
1st,5,CVE-2009-2561,1,0,100.00
1st,6,CVE-2009-4377,1,0,100.00
1st,7,CVE-2009-4376,1,0,100.00
1st,8,CVE-2010-2286,1,0,100.00
1st,9,CVE-2010-0304,1,0,100.00
1st,10,CVE-2010-2285,1,0,100.00
1st,11,CVE-2010-2284,1,0,100.00
1st,12,CVE-2010-2283,1,0,100.00
1st,13,CVE-2009-2559,1,0,100.00
1st,14,CVE-2009-3550,1,0,100.00
1st,15,CVE-2009-3549,1,0,100.00
```

Preliminary Results IX

```
1st,16,CVE-2010-1455,8,1,88.89
1st,17,CVE-2009-3243,3,1,75.00
1st,18,CVE-2009-3241,2,2,50.00
1st,19,CVE-2009-2560,1,1,50.00
1st,20,CVE-2009-3242,1,1,50.00
2nd,1,CVE-2009-3829,1,0,100.00
2nd,2,CVE-2009-2563,1,0,100.00
2nd,3,CVE-2009-2562,1,0,100.00
2nd,4,CVE-2009-4378,1,0,100.00
2nd,5,CVE-2009-2561,1,0,100.00
2nd,6,CVE-2009-4377,1,0,100.00
2nd,7,CVE-2009-4376,1,0,100.00
2nd,8,CVE-2010-2286,1,0,100.00
2nd,9,CVE-2010-0304,1,0,100.00
2nd,10,CVE-2010-2285,1,0,100.00
2nd,11,CVE-2010-2284,1,0,100.00
2nd,12,CVE-2010-2283,1,0,100.00
2nd,13,CVE-2009-2559,1,0,100.00
2nd,14,CVE-2009-3550,1,0,100.00
2nd,15,CVE-2009-3549,1,0,100.00
2nd,16,CVE-2010-1455,8,1,88.89
2nd,17,CVE-2009-3243,3,1,75.00
```

Preliminary Results X

```
2nd,18,CVE-2009-3241,3,1,75.00
2nd,19,CVE-2009-2560,1,1,50.00
2nd,20,CVE-2009-3242,1,1,50.00
```

Preliminary Results XI

Wireshark, CWE-based (NLP)

Notice ≈ 22% are in the 2nd guesses:

```
guess,run,config,good,bad,%
1st,1,-cweid -nopreprep -char -unigram -add-delta ,21,15,58.33
2nd,1,-cweid -nopreprep -char -unigram -add-delta ,29,7,80.56
guess,run,config,good,bad,%
1st,1,CWE-399,1,0,100.00
1st,2,CWE-189,1,0,100.00
1st,3,CWE-20,8,2,80.00
1st,4,NVD-CWE-Other,2,1,66.67
1st,5,NVD-CWE-noinfo,8,9,47.06
1st,6,CWE-119,1,3,25.00
2nd,1,CWE-399,1,0,100.00
2nd,2,CWE-189,1,0,100.00
2nd,3,CWE-20,9,1,90.00
2nd,4,NVD-CWE-Other,3,0,100.00
2nd,5,NVD-CWE-noinfo,13,4,76.47
2nd,6,CWE-119,2,2,50.00
```

Preliminary Results XII

Chrome, CVE-based

```
guess,run,config,good,bad,%
1st,1,-nopreprep -raw -fft -eucl ,10,1,90.91
1st,2,-nopreprep -raw -fft -cos ,10,1,90.91
1st,3,-nopreprep -raw -fft -diff ,10,1,90.91
1st,4,-nopreprep -raw -fft -cheb ,10,1,90.91
1st,5,-nopreprep -raw -fft -mink ,9,2,81.82
1st,6,-nopreprep -raw -fft -hamming ,9,2,81.82
2nd,1,-nopreprep -raw -fft -eucl ,11,0,100.00
2nd,2,-nopreprep -raw -fft -cos ,11,0,100.00
2nd,3,-nopreprep -raw -fft -diff ,11,0,100.00
2nd,4,-nopreprep -raw -fft -cheb ,11,0,100.00
2nd,5,-nopreprep -raw -fft -mink ,10,1,90.91
2nd,6,-nopreprep -raw -fft -hamming ,10,1,90.91
guess,run,config,good,bad,%
1st,1,CVE-2010-2301,6,0,100.00
1st,2,CVE-2010-2300,6,0,100.00
1st,3,CVE-2010-2299,6,0,100.00
1st,4,CVE-2010-2298,6,0,100.00
1st,5,CVE-2010-2297,6,0,100.00
```

Preliminary Results XIII

```
1st,6,CVE-2010-2304,6,0,100.00
1st,7,CVE-2010-2303,6,0,100.00
1st,8,CVE-2010-2295,10,2,83.33
1st,9,CVE-2010-2302,6,6,50.00
2nd,1,CVE-2010-2301,6,0,100.00
2nd,2,CVE-2010-2300,6,0,100.00
2nd,3,CVE-2010-2299,6,0,100.00
2nd,4,CVE-2010-2298,6,0,100.00
2nd,5,CVE-2010-2297,6,0,100.00
2nd,6,CVE-2010-2304,6,0,100.00
2nd,7,CVE-2010-2303,6,0,100.00
2nd,8,CVE-2010-2295,10,2,83.33
2nd,9,CVE-2010-2302,12,0,100.00
```

Preliminary Results XIV

Chrome, CWE-based

```
guess,run,config,good,bad,%
1st,1,-cweid -nopreprep -raw -fft -cheb ,9,0,100.00
1st,2,-cweid -nopreprep -raw -fft -cos ,9,0,100.00
1st,3,-cweid -nopreprep -raw -fft -diff ,9,0,100.00
1st,4,-cweid -nopreprep -raw -fft -eucl ,8,1,88.89
1st,5,-cweid -nopreprep -raw -fft -hamming ,8,1,88.89
1st,6,-cweid -nopreprep -raw -fft -mink ,6,3,66.67
2nd,1,-cweid -nopreprep -raw -fft -cheb ,9,0,100.00
2nd,2,-cweid -nopreprep -raw -fft -cos ,9,0,100.00
2nd,3,-cweid -nopreprep -raw -fft -diff ,9,0,100.00
2nd,4,-cweid -nopreprep -raw -fft -eucl ,8,1,88.89
2nd,5,-cweid -nopreprep -raw -fft -hamming ,8,1,88.89
2nd,6,-cweid -nopreprep -raw -fft -mink ,8,1,88.89
guess,run,config,good,bad,%
1st,1,CWE-79,6,0,100.00
1st,2,NVD-CWE-noinfo,6,0,100.00
1st,3,CWE-399,6,0,100.00
1st,4,CWE-119,6,0,100.00
1st,5,CWE-20,6,0,100.00
```

Preliminary Results XV

```
1st,6,NVD-CWE-Other,10,2,83.33
1st,7,CWE-94,9,3,75.00
2nd,1,CWE-79,6,0,100.00
2nd,2,NVD-CWE-noinfo,6,0,100.00
2nd,3,CWE-399,6,0,100.00
2nd,4,CWE-119,6,0,100.00
2nd,5,CWE-20,6,0,100.00
2nd,6,NVD-CWE-Other,11,1,91.67
2nd,7,CWE-94,10,2,83.33
```

Preliminary Results XVI

Tomcat, CVE-based

```
1st,1,-nopreprep -raw -fft -diff ,36,7,83.72
1st,2,-nopreprep -raw -fft -cheb ,36,7,83.72
1st,3,-nopreprep -raw -fft -cos ,37,9,80.43
1st,4,-nopreprep -raw -fft -eucl ,34,9,79.07
1st,5,-nopreprep -raw -fft -mink ,28,15,65.12
1st,6,-nopreprep -raw -fft -hamming ,26,17,60.47
2nd,1,-nopreprep -raw -fft -diff ,40,3,93.02
2nd,2,-nopreprep -raw -fft -cheb ,40,3,93.02
2nd,3,-nopreprep -raw -fft -cos ,40,6,86.96
2nd,4,-nopreprep -raw -fft -eucl ,36,7,83.72
2nd,5,-nopreprep -raw -fft -mink ,31,12,72.09
2nd,6,-nopreprep -raw -fft -hamming ,29,14,67.44
guess,run,config,good,bad,%
1st,1,CVE-2006-7197,6,0,100.00
1st,2,CVE-2006-7196,6,0,100.00
1st,3,CVE-2006-7195,6,0,100.00
1st,4,CVE-2009-0033,6,0,100.00
1st,5,CVE-2007-3386,6,0,100.00
1st,6,CVE-2009-2901,3,0,100.00
```

Preliminary Results XVII

```
1st,7,CVE-2007-3385,6,0,100.00
1st,8,CVE-2008-2938,6,0,100.00
1st,9,CVE-2007-3382,6,0,100.00
1st,10,CVE-2007-5461,6,0,100.00
1st,11,CVE-2007-6286,6,0,100.00
1st,12,CVE-2007-1858,6,0,100.00
1st,13,CVE-2008-0128,6,0,100.00
1st,14,CVE-2007-2450,6,0,100.00
1st,15,CVE-2009-3548,6,0,100.00
1st,16,CVE-2009-0580,6,0,100.00
1st,17,CVE-2007-1355,6,0,100.00
1st,18,CVE-2008-2370,6,0,100.00
1st,19,CVE-2008-4308,6,0,100.00
1st,20,CVE-2007-5342,6,0,100.00
1st,21,CVE-2008-5515,19,5,79.17
1st,22,CVE-2009-0783,11,4,73.33
1st,23,CVE-2008-1232,13,5,72.22
1st,24,CVE-2008-5519,6,6,50.00
1st,25,CVE-2007-5333,6,6,50.00
1st,26,CVE-2008-1947,6,6,50.00
1st,27,CVE-2009-0781,6,6,50.00
1st,28,CVE-2007-0450,5,7,41.67
```

Preliminary Results XVIII

```
1st,29,CVE-2007-2449,6,12,33.33
1st,30,CVE-2009-2693,2,6,25.00
1st,31,CVE-2009-2902,0,1,0.00
2nd,1,CVE-2006-7197,6,0,100.00
2nd,2,CVE-2006-7196,6,0,100.00
2nd,3,CVE-2006-7195,6,0,100.00
2nd,4,CVE-2009-0033,6,0,100.00
2nd,5,CVE-2007-3386,6,0,100.00
2nd,6,CVE-2009-2901,3,0,100.00
2nd,7,CVE-2007-3385,6,0,100.00
2nd,8,CVE-2008-2938,6,0,100.00
2nd,9,CVE-2007-3382,6,0,100.00
2nd,10,CVE-2007-5461,6,0,100.00
2nd,11,CVE-2007-6286,6,0,100.00
2nd,12,CVE-2007-1858,6,0,100.00
2nd,13,CVE-2008-0128,6,0,100.00
2nd,14,CVE-2007-2450,6,0,100.00
2nd,15,CVE-2009-3548,6,0,100.00
2nd,16,CVE-2009-0580,6,0,100.00
2nd,17,CVE-2007-1355,6,0,100.00
2nd,18,CVE-2008-2370,6,0,100.00
2nd,19,CVE-2008-4308,6,0,100.00
```

Preliminary Results XIX

```
2nd,20,CVE-2007-5342,6,0,100.00
2nd,21,CVE-2008-5515,19,5,79.17
2nd,22,CVE-2009-0783,12,3,80.00
2nd,23,CVE-2008-1232,13,5,72.22
2nd,24,CVE-2008-5519,12,0,100.00
2nd,25,CVE-2007-5333,6,6,50.00
2nd,26,CVE-2008-1947,6,6,50.00
2nd,27,CVE-2009-0781,12,0,100.00
2nd,28,CVE-2007-0450,7,5,58.33
2nd,29,CVE-2007-2449,8,10,44.44
2nd,30,CVE-2009-2693,4,4,50.00
2nd,31,CVE-2009-2902,0,1,0.00
```

Preliminary Results XX

Tomcat, CWE-based

```
guess,run,config,good,bad,%
1st,1,-cweid -nopreprep -raw -fft -cheb ,27,6,81.82
1st,2,-cweid -nopreprep -raw -fft -diff ,27,6,81.82
1st,3,-cweid -nopreprep -raw -fft -cos ,24,9,72.73
1st,4,-cweid -nopreprep -raw -fft -eucl ,13,20,39.39
1st,5,-cweid -nopreprep -raw -fft -hamming ,12,21,36.36
1st,6,-cweid -nopreprep -raw -fft -mink ,9,24,27.27
2nd,1,-cweid -nopreprep -raw -fft -cheb ,32,1,96.97
2nd,2,-cweid -nopreprep -raw -fft -diff ,32,1,96.97
2nd,3,-cweid -nopreprep -raw -fft -cos ,29,4,87.88
2nd,4,-cweid -nopreprep -raw -fft -eucl ,17,16,51.52
2nd,5,-cweid -nopreprep -raw -fft -hamming ,18,15,54.55
2nd,6,-cweid -nopreprep -raw -fft -mink ,13,20,39.39
guess,run,config,good,bad,%
1st,1,CWE-264,7,0,100.00
1st,2,CWE-255,6,0,100.00
1st,3,CWE-16,6,0,100.00
1st,4,CWE-119,6,0,100.00
1st,5,CWE-20,6,0,100.00
```

Preliminary Results XXI

```
1st,6,CWE-200,22,4,84.62
1st,7,CWE-79,24,21,53.33
1st,8,CWE-22,35,61,36.46
2nd,1,CWE-264,7,0,100.00
2nd,2,CWE-255,6,0,100.00
2nd,3,CWE-16,6,0,100.00
2nd,4,CWE-119,6,0,100.00
2nd,5,CWE-20,6,0,100.00
2nd,6,CWE-200,23,3,88.46
2nd,7,CWE-79,30,15,66.67
2nd,8,CWE-22,57,39,59.38
```

Preliminary Results XXII

Typical output fragment

```
                 File: wireshark-1.2.0/epan/dissectors/packet-afs.c
               Config: -nopreprep -raw -fft -cheb -graph
      Processing time: 0d:0h:0m:0s:156ms:156ms
         Subject's ID: 20092562
   Subject identified: CVE-2009-2562
...
Expected subject's ID: 20092562 (possible: [20092562])
     Expected subject: CVE-2009-2562
       Second Best ID: 3
     Second Best Name: CVE-2010-2285
           Date/time: Fri Oct 01 13:48:09 EDT 2010
```

Shortcomings

- Looking at a signal is less intuitive visually for code analysis.
- Line numbers! (easily “filtered out” as high-frequency “noise”, etc.). A whole “relativistic” and machine learning methodology developed for the line numbers.
- Accuracy depends on the quality of the knowledge base. “Garbage in – garbage out.”
- To detect CVE or CWE signatures in non-CVE cases requires large knowledge bases (human-intensive to collect).
- No path tracing (since no parsing is present); no slicing, semantic annotations, context, locality of reference, etc.
- Lots of algorithms and their combinations to try (currently ≈ 1800 permutations).

Advantages

- Relatively fast (e.g. Wireshark $\approx$ 2400 files train and test about 3 minutes)
- Language independent (no parsing) – given enough examples can apply to any language, i.e. methodology is the same no matter C, C++, Java or any other source or binary language.
- Can automatically learn a large knowledge base to test on known and unknown cases.
- Can be used to quickly pre-scan projects for further analysis by humans and other tools.
- Can learn from other SATE'10 reports.
- Can learn from SATE'09 and SATE'08 reports.
- High precision in CVEs and CWE detection.
- Lots of algorithms and their combinations.

Conclusion

Practical implications:

- The approach can be used on any target language without modifications to the methodology or knowing the syntax of the language.
- The approach can nearly identically be transposed onto the compiled binaries and bytecode, detecting vulnerable deployments and installations – sort of like virus scanning of binaries, but instead scanning for security-weak binaries on site deployments to alert sysadmins.
- Can learn from binary signatures from other tools like Snort.
- Open-source MARF already is; MARFCAT will be published soon after the workshop along with the e-print documentation in [Mok10c].

Questions?

Thank you :-)

References I

- Serguei A. Mokhov and Mourad Debbabi.
  File type analysis using signal processing techniques and machine learning vs. `file` unix utility for forensic analysis.
  In Oliver Goebel, Sandra Frings, Detlef Guenther, Jens Nedon, and Dirk Schadt, editors, *Proceedings of the IT Incident Management and IT Forensics (IMF'08)*, pages 73–85, Mannheim, Germany, September 2008. GI.
  LNI140.

- Serguei A. Mokhov, Marc-André Laverdière, and Djamel Benredjem.
  Taxonomy of linux kernel vulnerability solutions.
  In *Innovative Techniques in Instruction Technology, E-learning, E-assessment, and Education*, pages 485–493, University of Bridgeport, U.S.A., 2007.
  Proceedings of CISSE/SCSS'07.

- Serguei A. Mokhov.
  Study of best algorithm combinations for speech processing tasks in machine learning using median vs. mean clusters in MARF.
  In Bipin C. Desai, editor, *Proceedings of C3S2E'08*, pages 29–43, Montreal, Quebec, Canada, May 2008. ACM.
  ISBN 978-1-60558-101-9.

References II

Serguei A. Mokhov.
Complete complimentary results report of the MARF's NLP approach to the DEFT 2010 competition.
[online], June 2010.
http://arxiv.org/abs/1006.3787.

Serguei A. Mokhov.
L'approche MARF à DEFT 2010: A MARF approach to DEFT 2010.
In *Proceedings of TALN'10*, July 2010.
To appear in DEFT 2010 System competition at TALN 2010.

Serguei A. Mokhov.
The use of machine learning with signal- and NLP processing of source code to detect and classify vulnerabilities and weaknesses with MARFCAT.
[online], October 2010.
To appear; online at http://arxiv.org/abs/1010.2511.

NIST.
National Vulnerability Database.
[online], 2005–2010.
http://nvd.nist.gov/.

References III

NIST.
National Vulnerability Database statistics.
[online], 2005–2010.
http://web.nvd.nist.gov/view/vuln/statistics.

Vadim Okun, Aurelien Delaitre, Paul E. Black, and NIST SAMATE.
Static Analysis Tool Exposition (SATE) 2010.
[online], 2010.
See http://samate.nist.gov/SATE.html and
http://samate.nist.gov/SATE2010Workshop.html.

The MARF Research and Development Group.
The Modular Audio Recognition Framework and its Applications.
[online], 2002–2010.
http://marf.sf.net and http://arxiv.org/abs/0905.1235, last viewed April 2010.

Various contributors and MITRE.
Common Weakness Enumeration (CWE) – a community-developed dictionary of software weakness types.
[online], 2010.
See http://cwe.mitre.org.